STATISTIQUES À DEUX VARIABLES

I. Rappels de la 1ère STG

1. Population

Définition : On appelle **population**, l'ensemble étudié dans une étude statistique.

Exemples : La classe de la T le STGM du lycée Marguerite Yourcenar (91), les **planètes** du système solaire.

Définition : On appelle **individu**, tout élément de la population étudiée.

Exemples :

- Bernaud, Coursière, Deminiac et Leal sont des « individus » si la population étudiée est la T le STGM.
- Mars, Jupiter, Sature, Terre sont des « individus » si la population étudiée est l'ensemble des planètes du système solaire.

Définition : On appelle échantillon de taille *n*, tout groupe de *n* individus pris dans la population.

- Le groupe {De Waard; Le Bouill ; Marques} est un échantillon de taille 3 de la T le STGM.
- Le groupe {Mercure, Vénus, Terre, Mars, Jupiter, Saturne, Uranus} est un échantillon de taille 7 de la population des planètes du système solaire.

2. Variable statistique ou caractère

Définition : On appelle **caractère** ou **variable statistique**, l'aspect étudié sur chaque individu d'une population.

Exemples :

- Si on demande à chaque élève, leur taille, le caractère étudié est donc la taille.
- Si on demande à chaque élève, leur note au bac blanc (février 2010), le caractère étudié est donc leur note.
- Pour les planètes du système solaire, le caractère peut être la taille, la distance par rapport au soleil, la distance par rapport à la terre, etc.
- Pour un pays, le caractère peut être la superficie, la population, le P.I.B., etc …

Définition : La **valeur du caractère** est le résultat possible, la mesure possible ou la réponse possible d'une observation du caractère.

Il y a deux types de caractère : le caractère quantitatif et le caractère qualitatif.

Définition : On dit qu'un **caractère est quantitatif**, lorsque la valeur du caractère peut être mesurée (il est donc exprimé grâce à un nombre).

Exemples : La taille, le poids, l'âge, le nombre d'habitants, le nombre de voix d'un candidat à une élection, etc.

Définition : On dit qu'un **caractère est qualitatif**, lorsque la valeur du caractère ne peut pas être mesurée.

Exemples : La couleur préférée, le vote, le moyen de locomotion pour aller au lycée, le prénom, etc.

Remarque : Désormais, les caractères étudiés seront toujours **quantitatif**.

II. Série statistique à deux variables

Définition :

Une série statistique double est une liste de la forme $(x_1 ; y_1)$; $(x_2 ; y_2)$; … ; $(x_n ; y_n)$
où $(x_1; y_1)$; $(x_2; y_2)$; ...; $(x_n; y_n)$ correspondent aux données concernant **deux** caractères étudiés sur chacun des n individus de la population.

Ces séries sont le plus souvent présentées sous forme de tel tableau :

x_i	x_1	x_2		x_n
y_i	y_1	y_2		y_n

Exemples :

*

Durée (en heures par semaine)	1	2	3	4	5	6
Montant (en euros)	135	230	320	405	4075	520

* La taille et l'âge (en année) de chaque élève du lycée Marguerite Yourcenar en 2009 – 2010.

* Un chef d'entreprise a fait un relevé sur cinq années de l'évolution du pourcentage d'emplois à temps partiel dans son entreprise :

Année	2001	2002	2003	2004	2005
Rang (x_i)	1	2	3	4	5
Pourcentages d'emplois partiels (y_i)	6,5	12,5	16,9	20,6	23,5

* La population et le Produit Intérieur Brut des pays de la communauté européenne.

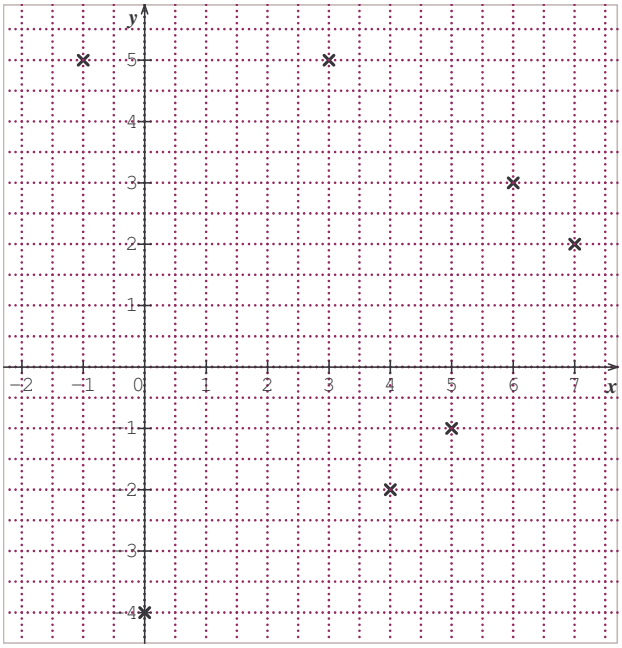
II. Représentation graphique

Définition : Représentation graphique d'une série statistique double

Soit $(x_1 ; y_1)$; $(x_2 ; y_2)$; … ; $(x_n ; y_n)$ une série statistique double.

La représentation graphique de cette série est **l'ensemble des *n* points** que l'on peut noter A_1 ; A_2 ; … ; A_n tel que :
A_1 ait pour coordonnées $(x_1 ; y_1)$ (ce qui signifie que l'abscisse de A_1 est x_1 et que l'ordonnée de A_1 est y_1).
A_2 ait pour coordonnées $(x_2 ; y_2)$ (ce qui signifie que l'abscisse de A_2 est x_2 et que l'ordonnée de A_2 est y_2).
…
A_n ait pour coordonnées $(x_n ; y_n)$ (ce qui signifie que l'abscisse de A_n est x_n et que l'ordonnée de A_n est y_n).

L'ensemble des points A_1 ; A_2 ; … ; A_n est aussi appelé, « ***nuage de points*** » associé à la série statistique $(x_1 ; y_1)$; $(x_2 ; y_2)$; … ; $(x_n ; y_n)$.



Exemple :

Soit la série statistique double suivante :

x_i	– 1	0	1	2	3
y_i	5	– 4	8	8	5

x_i	4	5	6	7
y_i	– 2	– 1	3	2

On a donc $\boldsymbol{n = 9}$
$x_1 = -1$; $x_2 = 0$; $x_3 = 1$; $x_4 = 2$; … ; $x_8 = 6$ et $x_9 = 7$.

$y_1 = 5$; $y_2 = -4$; $y_3 = 8$; $y_4 = 8$; ... ; $y_8 = 3$ et $y_9 = 2$.

Le point A_1 a donc pour abscisse $x_1 = -1$ et pour ordonnée $y_1 = 5$

Le point A_2 a donc pour abscisse $x_2 = 0$ et pour ordonnée $y_2 = -4$

Le point A_3 a donc pour abscisse $x_3 = 1$ et pour ordonnée $y_3 = 8$

Etc …

Le point A_9 a donc pour abscisse $x_9 = 7$ et pour ordonnée $y_9 = 2$

On obtient le nuage de point suivant (page précédente à droite du tableau) :

Définition : Point moyen d'un nuage de point

Soit $A_1(x_1 ; y_1)$; $A_2(x_2 ; y_2)$; … ; $A_n(x_n ; y_n)$ un nuage de points associé à la série statistique double.

x_i	x_1	x_2		x_n
y_i	y_1	y_2		y_n

Le point moyen G de ce nuage est le point qui a :

- Pour **abscisse**, la moyenne des abscisses c'est-à-dire : $\dfrac{x_1 + x_2 + \dots x_n}{n}$.

- Pour **ordonnée**, la moyenne des ordonnées, c'est-à-dire : $\dfrac{y_1 + y_2 + \dots y_n}{n}$.

Exemple :

Soit le nuage de points précédent associé à la série statistique suivante :

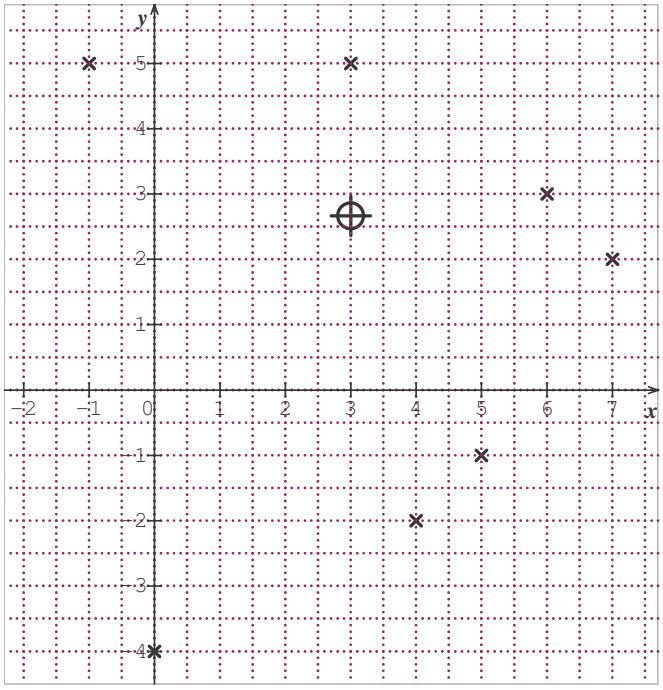
x_i	– 1	0	1	2	3	4	5	6	7
y_i	5	– 4	8	8	5	– 2	– 1	3	2

Le point moyen G a donc pour abscisse :

$$\frac{-1 + 0 + 1 + 2 + 3 + 4 + 5 + 6 + 7}{9} = \mathbf{3}$$

Le point moyen G a donc pour ordonnée :

$$\frac{5 + (-4) + 8 + 8 + 5 + (-2) + (-1) + 3 + 2}{9} \approx \mathbf{2{,}67.}$$



Remarque :

* Dans la pratique, une fois les coordonnées calculées, on placera le point selon la précision que permettra le repère.

* Le logiciel graphique que j'ai utilisé désigne le point moyen par une cible, mais vous **devez le représenter comme un point normal.**

III. Ajustement linéaire

1. Définition : Ajustement linéaire

Définition :

Soit $A_1(x_1 ; y_1)$; $A_2(x_2 ; y_2)$; … ; $A_n(x_n ; y_n)$ un nuage de points associé à la série statistique double.

Un ajustement linéaire de ce nuage est la droite qui « passe le plus près » des points du nuage.

Remarque : L'ajustement linéaire se nomme aussi « **ajustement affine** » ou « **droite de régression** ».

2. Méthodes d'ajustement linéaire

Ces méthodes n'ont de pertinence que si l'allure du nuage de points est « globalement rectiligne ».
Cela signifie que les points sont assez bien regroupés autours d'une droite.

Par exemple, l'ajustement linéaire n'est pas pertinent pour le nuage de point précédent.

2.1 Ajustement linéaire par la méthode de Mayer :

a) On découpe le nuages de points en deux sous nuages de quantité identique (ou à 1 près si le nombre de points est impair).

b) On détermine G_1, le ***point moyen*** du sous nuage 1.

c) On détermine G_2, le ***point moyen*** du sous nuage 2.

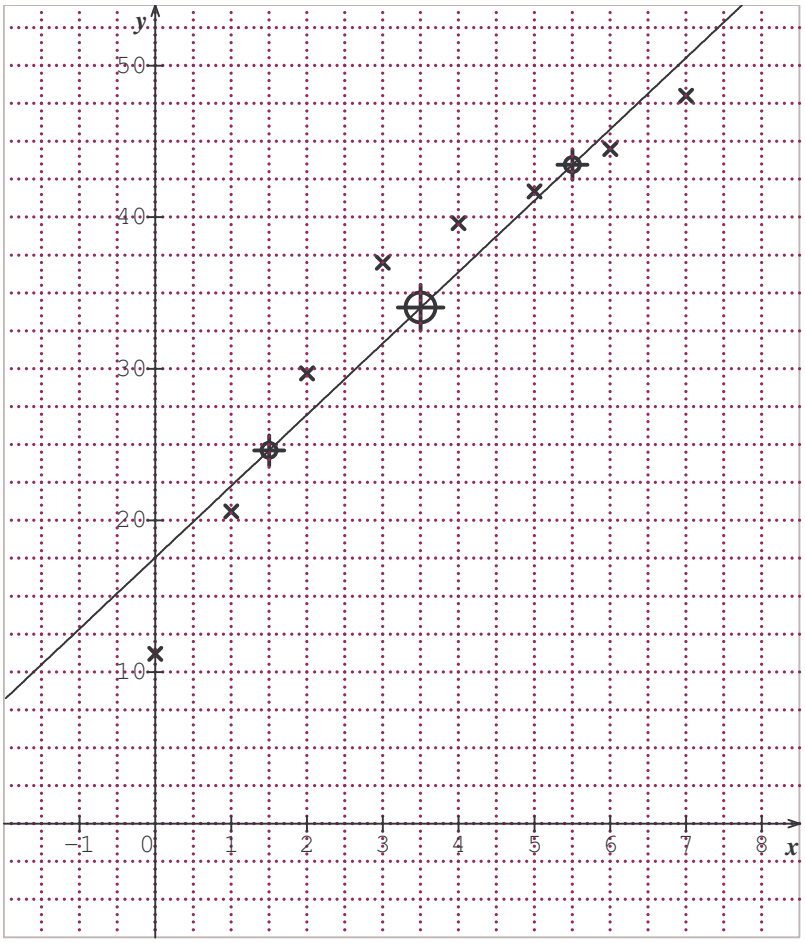
d) On trace la droite (G_1 G_2) qui est l'ajustement affine recherché.

Exemple :

Soit la série statistique double suivante :

x_i	0	1	2	3
y_i	11,2	20,6	29,7	37,0

x_i	4	5	6	7
y_i	39,6	41,7	44,5	48,0



Les deux petites cibles représentent les points moyens G_1 et G_2 et la grande cible désigne le point moyen G de tous les 8 points.

a) On a donc découpé le nuage de points en deux sous nuages de 4 points chacun.

b) G_1 est le point moyen des points de coordonnées respectives

(0 ; 11,2) ; (1 ; 20,6) ; (2 ; 29,7) et (3 ; 39,6).

c) G_2 est le point moyen des points de coordonnées respectives

(4 ; 39,6) ; (5 ; 41,7) ; (6; 44,5) et (7 ; 48,0).

d) On trace la droite (G_1 G_2).

2.2 Ajustement linéaire par la méthode des moindres carrés (calculatrice) :

On détermine cet ajustement affine grâce à une fonction spéciale de la calculatrice.

Vous devez consulter le mode d'emploi de votre calculatrice pour apprendre à rentrer une série statistique double et à déterminer la droite de régression linéaire.

Dans le livre utilisé lors de cette année, vous trouverez (couverture intérieure au début et à la fin) comment éditer (rentrer dans la calculatrice) une série statistique double.

Sinon, voilà deux autres modes d'emploi pour les calculatrices de type ***Casio*** et ***Texas-Instrument***.

* mode d'emploi des calculatrices ***TI*** pour les statistiques. http://www.merkiled.fr/TSTGM/calc_stg1.pdf

* mode d'emploi des calculatrices ***casio*** pour les statistiques. http://www.merkiled.fr/TSTGM/calc_stg2.pdf